

הפרויקט : אפיון והצעת חלופות בנוגע לפערים במערכת החינוך בהישגים ובזכאות לבגרות

1. תיאור הבעיה :

מערכת החינוך מהווה נדבך מרכזי בעיצוב העתיד של צעירי ישראל, ולהצלחה במערכת זו השפעה ניכרת על חייהם הבוגרים (קבלה לתפקידים משמעותיים בצבא, השכלה גבוהה, כושר השתכרות...). על אף שהמערכת מקבלת תקצוב לכל תלמיד, קיימים פערים בתקצוב ובשירותי החינוך באזורים שונים בארץ. אנו מעריכים כי ניתן למצוא דפוסים במידע שיאפשרו איתור פערי מדיניות או אתגרים שיש להתמקד בהם - ע"ב השוואה בין פרטי מידע שונים על תקציב, שעות לימוד, ומאפייני המסגרת והדמוגרפיה של בתי הספר (ממלכתי / דתי / חרדי, אזור מגורים ונפה, חתך סוציאקונומי, מגזר וכו').

2. תיאור של הדאטה :

- א. 13 קבצי אקסל הכוללים חיתוכים שונים מהאתר של משרד החינוך ([היפר-קישור](#)):
גודל הקבצים - סה"כ MB16.2, טבלאות אשר נגזרו ככל הנראה ממקור מרכזי אחד (או שניים לכל היותר), המכילים נתוני תפעול בתי הספר (תקציב, שעות חינוך, היקפי כיתות וכו') ושל תוצאות במדדים חיצוניים (מבחני מיצ"ב, תוצאות בגרות, אחוזי גיוס לצבא ולשירות לאומי). לאחר ניתוח הנתונים התמקדנו בטבלאות זכאות לבגרות ותקציב בתי הספר.
- ב. קובץ אקסל אחד של נתוני זכאות לבגרות 2013-2016 ([היפר-קישור](#)):
גודל הקובץ - KB8860, טבלה בת 23896 שורות, כל שורה מייצגת נתונים של בית ספר באחת השנים הנ"ל. בקובץ 53 עמודות המייצגות נתונים שונים על בית הספר.
- ג. נתונים רשמיים ממשרד החינוך על מיקומי בתי ספר : קובץ אקסל המכיל מידע רשמי של משרד החינוך על בתי הספר, אליו נוספו מיקומים גיאוגרפיים לפי כתובת של כל בית ספר. בגודל KB356. בקובץ 21 עמודות המייצגות נתונים שונים של בתי הספר, על כ-2089 שורות (כל שורה - בית ספר). את הקובץ קיבלנו באדיבותו של חברנו לספסל הלימודים - עציון הררי.

3. הפתרון שלנו :

אנו התמקדנו בפתרונות של ניתוח ואגרגציה של הנתונים, והנגשה שלהם כתוצרים ניתנים להבנה ולשימוש. מטרתנו לבדוק את תוצאות מערכת החינוך בזכאות לבגרות למול מאפייני תקציב, פיקוח ושירותי חינוך נוספים לבתי ספר ממגזרים ומעמד סוציאקונומי שונים. זאת במטרה לאפשר ערעור על פרדיגמות קיימות ויצירת עזרים לגיבוש מדיניות חדשה מבוססת נתונים. בהתאם, התוצר שלנו מורכב משני חלקים :

1. ניתוח וסקירת מאקרו מעמיקה על התפלגויות בתי הספר לפי שנות הקמה, חלוקה למאפיינים דמוגרפיים שונים, ניתוח היבטים תקציביים שלהם, וניתוח כללי של זכאויות לבגרות והפערים ביניהם. זאת במטרה הן להבין את המידע והן כתיאור כללי של מערכת החינוך "ממעוף הציפור".
 2. שימוש במספר מודלים לתיאור וניבוי זכאות לבגרות לפי יתר מאפייני בתי הספר, והשוואתם לחלוקות הקיימות של בתי הספר ל"קבוצות דומים" - בכך:
 1. חלוקה לתתי קבוצות לפי שילובי מאפיינים חדש (מגזר ואחוזון טיפוח).
 2. קלאסטרים המבוססים על כלל נתוני הזכאות לבגרות.
 3. עץ החלטה לניתוח התפלגות בתי הספר לפי מאפיין נמדד.
- זאת תוך הנגשה של התוצרים כמידע טבלאי, תוצרים ויזואליים וממשקים אינטראקטיביים - כשכבות על מפה, וכממשק ליצירת עץ החלטות.
- התבססנו על המודלים הקיימים בספריות "pandas", "scikit learn", "geopy", "numpy", ו-"matplotlib" ושירותים של GOOGLE MAPS תוך התאמה שלהם לצרכינו :
- מודל K-Means בגדלים המתאימים למידע (ע"ב ניתוח Silhouette SSE)
 - מודל עץ החלטות - תוך הגבלת עומק, לאור הצורך בקריאות ואפשרות של היעדר הגעה לסיום עם סינגלטונים.
 - ניתוח מילים ליצירת גרפים וענני מילים (בעזרת אתר <https://www.wordclouds.com/>)
 - שימוש ב MY MAPS, שירות של GOOGLE MAPS על מנת להעלות את הנתונים המעובדים עבור ערים ובתי ספר על מפה.
- בהתאם לאמור, יצרנו מודלים שונים של הערכת אמינות התוצאות :
1. בחינת נתוני הניתוח למול נתונים גולמים בחלוקה לערים ורשויות.
 2. שילוב של דגימה והשוואה ידנית של מקרי בוחן מהתוצרים לבחינת מציאות.
 3. בחנו את הנחות המוצא שלנו למול גורמים מהתחום (קרובי משפחה העובדים במשרד החינוך) וכן ע"ב חומרי רקע של משרד החינוך שסקרו את מאפייני המדדים ונתוני סטטוס שונים.
 4. כמו כן השתמשנו בהשוואות שונות בין המודלים (עליהן נרחיב בפרק התוצרים על כל מודל) כדי לבדוק את תוצאות הניתוחים השונים והקוהרנטיות ביניהם.

פרק תוצרים

תוצרי ניתוח וסקירת מאקרו על בתי הספר:

*יש לציין כי בכל קבוצי הקוד נעזרו בספריית pandas לניתוח וספריית matplotlib ליצירת גרפים * לכל סעיף בתוצרים מצורפת תיקייה עם התוצרים עצמם.

1. שמות בתי ספר כמאפיין בסיסי לניתוח:

schoolNames.py: קובץ קוד: a

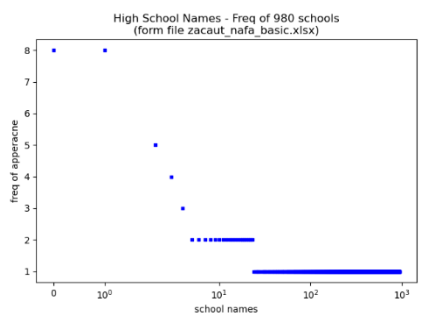
b. **מטרה:** אפיון כללי של בתי הספר בארץ לפי שמם – המהווה את פריט הידע הבסיסי ביותר הנתון לאדם המתעניין בבית הספר.

והתייחסנו לכל מילה כ'טוקן' בכוונה לאתר תבניות הקשורות לשמות בתי הספר. לאחר מכן יצרנו מילונים עם המילה וכמות ההופעות שלה ולשם ויזואליזציה נעזרנו באתר <https://www.wordclouds.com>

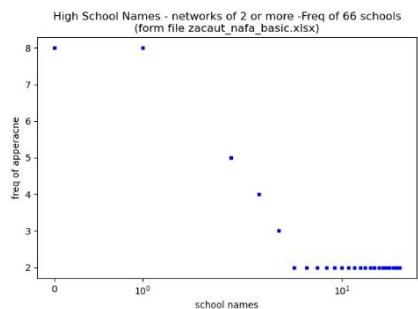
d. **בדיקת מהימנות:** היות והשמות המלאים אינם חד-חד ערכיים, אנו משתמשים בסמל מוסד להבחנה בין בתי הספר.

e. תוצרים ומסקנות:

i. מבחינה של כלל שמות בתי הספר בולטים המון שמות ייחודיים, ובמקרים שאינם כאלו – מדובר ברשות או במילים פוסלות (חינוך, ביי"ס וכו') או מאפיינות (לדוגמא ציבור דתי "בנע", "דתי", "אמית").

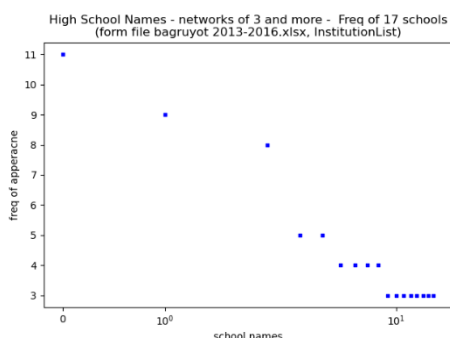


ii. מהרשימה שקיבלנו בשקף קודם, הורדנו את כל בתי ספר עם שם יחיד כדי לקבל 'רשתות' (רשת: שני בתי ספר ומעלה בעלי אותו השם) ואיתרנו 66 בתי ספר המהווים כ-24 רשתות. מעניין לראות ריבוי שמות עם מאפיינים דתיים (מבדיקה ידנית כעשרה מתוך 24 שמות יחודיים).



iii. לאור ריבוי השמות עם מאפיינים דתיים, בחינה

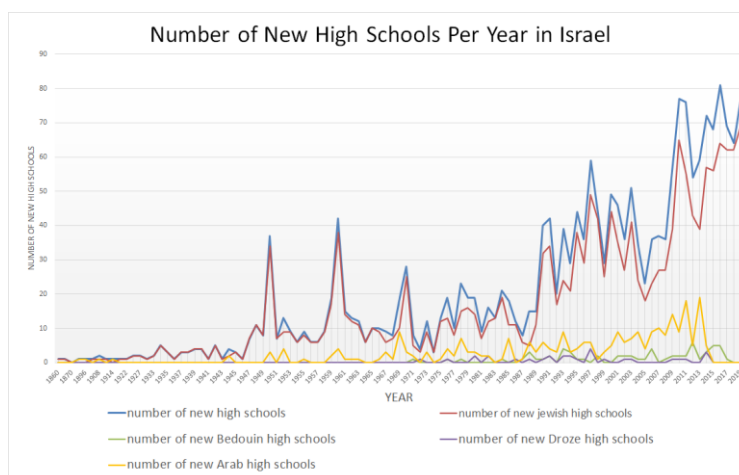
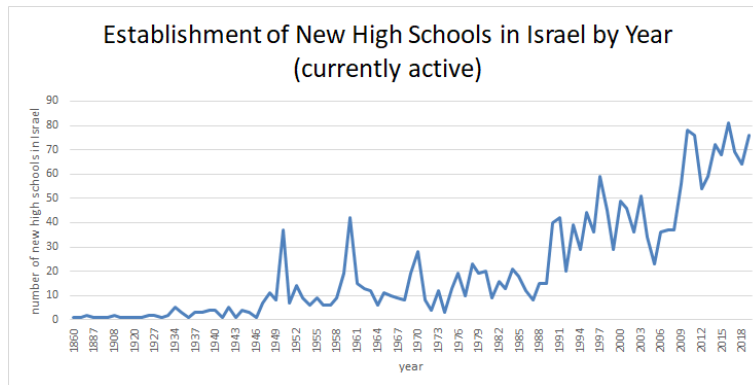
דומה של קובץ המכיל גם בתי ספר שאינם זכאים לבגרות או שאין לנו עליהם מידע מלא מעלה תמונה דומה (אם כי יותר ברורה, כאשר הרוב המוחלט של הרשתות הן דתיות). במקרה זה צימצום לרשתות בעלי 3 בתי ספר ומעלה העלה יחס אפילו יותר ברור של 8 בתי ספר דתיים מתוך 17, מה שעשוי להעיד על בולטות של רשתות חינוך בסקטור הדתי והחרדי.

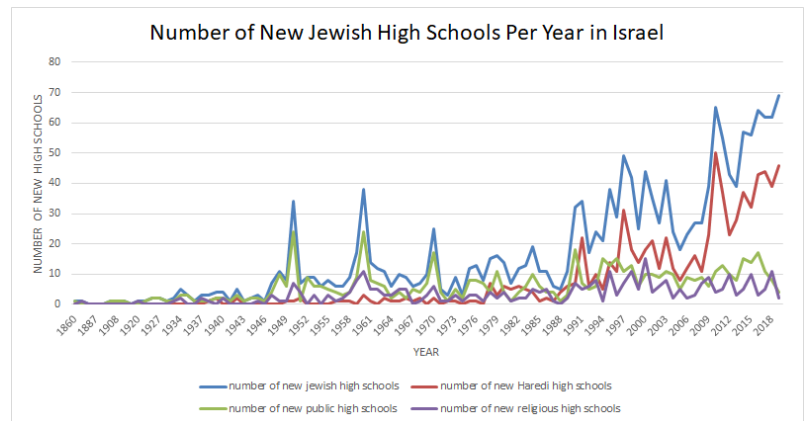


2. ניתוח פתיחת בתי ספר חדשים

- a. **מטרה:** אפיון התפלגות מועדי פתיחת בתי הספר כמדד לקיבעון או צמיחה של החינוך בסוגי הפיקוח ובמגזרים השונים.
- b. **הסבר התיאור:** ניתחנו את התפלגות הנתונים לאורך השנים ובחנו אותה בהתעמקות לפי מאפיינים.
- c. **בדיקת מהימנות:** לאור ההעמקה בתתי קבוצות, בחנו את התוצאות למול ידע על ההיסטוריה של החברה הישראלית.
- d. **תוצרים ומסקנות** (שליפת מידע ראשונה בקובץ newSchool.py. תוצר מלא ניתן לראות בקובץ new schools by years.xlsx):

- i. כלל בתי הספר לפי שנה – ניתן לראות עליה חדה בשנה הראשונה להקמת המדינה, או לדוגמא בשנים 1958-1964 עליה חדה (מתאים למספרים הרשמיים של העליות באותה תקופה, תוספת של 3000 אלף עולים בנוסף לגידול הטבעי), גרף ראשון.
- ii. פילוחים לפי שנה, שנה ומגזר – בהם ניתן לראות דומיננטיות ברורה של המגזר היהודי בהשפעה כמותנית על מגמות פתיחת בתי הספר לאורך השנים (בגרף השני: העקום הכחול – הכללי לעומת האדום – היהודי).
- iii. מבחינה ממוקדת במגזר היהודי בחלוקה לפי סוגי פיקוח (ממלכתי / ממלכתי דתי / חרדי), גרף שלישי, ניתן לשים לב כי עד שנות השמונים, רוב בתי הספר שנפתחו השתייכו לזרם הממלכתי (כלומר אוכלוסייה חילונית, עקומה ירוקה) עם השפעות של הזרם הממלכתי-דתי, ומתחילת שנות התשעים ישנה מגמה ברורה של התחזקות הזרם החרדי, כאשר מספר בתי הספר הנפתחים לטובת ציבור זה גדול באופן משמעותי מכל זרם אחר במגזר היהודי. מעניין לראות כי במקביל מספר התיכונים החדשים הנפתחים לטובת הציבור החילוני נמצא במגמת עלייה איטית, ולעיתים אף בירידה (בהתחשב בגידול דמוגרפי טבעי).
- iv. ישנם מספר הסברים לתופעה זאת. הראשון הוא גידול דמוגרפי מהיר של הזרם החרדי ביחס לאוכלוסייה הכללית, הדורש בהתאם יותר מוסדות לימוד. הסבר נוסף הוא המגמה של הממשלה בשני העשורים האחרונים לשלב לימודי ליבה בגילאי התיכון של הזרם החרדי. עניין זה דרש פתיחת בתי ספר חדשים המותאמים ללימודי קודש וחול, במתכונת מתאימה לציבור זה. בהתבוננות נוספת בנתונים שאספנו זיהינו כי מספר לא מבוטל של בתי ספר תיכונים המשתייכים לזרם החרדי מכילים מספר קטן של תלמידים (פחות מ-100) ביחס לבתי ספר המשתייכים לזרם החילוני, מה שכנראה משפיע על הגידול בכמות בתי הספר.





3. זכאות לבגרות לפי ערים ונפות גאוגרפיות:

a. **קובץ קוד:** `zacaut2.py` מורכב מסקריפטים שקיבצנו יחד לשם נוחות, בפועל כל חלק (מסומן בדוקומנטציה) רץ בנפרד.

b. **מטרה:** אפיון התפלגות הזכאות לבגרות בראי נתונים גיאוגרפיים, כמוקד משפיע פוטנציאלי, וכביסס לתוצר כשכבת מפה שתאפשר אקספלורציה, גמישות למשתמש והשוואה מבוססת מיקום.

c. **הסבר התהליך:**

i. **עבור הנפות:** איחדנו בין מאגרי המידע על זכאות בתי ספר (`school - bagrut.xlsx`) ועל שייכות לנפה (`schools_utm_only_hs.xlsx`) וחישובנו ממוצע מסוכם לכל נפה, בכל שנה ובממוצע על כל השנים הנתונות. קובץ סופי - `zacaut_per_nafa_mean.xlsx`

ii. **עבור הערים:** יצרנו קובץ עזר - `city_mean_zacaut_with_nafa.xlsx` בו חישובנו את ממוצע זכאות לכל עיר וגם הצמדנו לכל עיר את הנפה המתאימה לחישובים עתידיים. יצרנו קובץ עזר נוסף - בו הוספנו את החישוב של המיקום הגיאוגרפי - `City_mean_zacaut_geo_with_nafa.xlsx` - וחישובנו את המיקום הגיאוגרפי של כל יישוב/עיר על ידי חישוב ממוצע של המיקומים של בתי הספר הנתונים.

התמודדנו עם פערים בהשוואה (חוסר קונסיסטנטיות בשמות של כל המועצות אזוריות בארץ וערים עם שמות המורכבים ממספר מילים) באמצעות תיוג מקום ע"ב אתר <https://geohack.toolforge.org/> ולפי הרישום בוויקיפדיה, לצד השלמת פערים ידנית נקודתית. ליצירת הקובץ הסופי - `city_mean_zacaut_geo_with_nafa_clean.xlsx`

d. **בדיקת מהימנות:**

- בחינה למול נתונים כללים המתפרסמים בתקשורת – דוגמת העובדה כי ביהודה ושומרון (לפי אתרי חדשות) התקצוב לכל תלמיד הוא הגבוה ביותר בארץ, ואכן ניתן לראות כי באופן עקבי אחוז הזכאות בנפה זאת גבוהה מאוד ולרוב מקסימלית ברמה הארצית.

- מצדו השני של המתרס, באר שבע, צפת ואזור כנרת נפות עם אחוז זכאות נמוך יותר. לדוגמא נפת באר שבע מתפרשת על חצי משטחה של מדינת ישראל וכוללת אזורי פריפריה ואוכלוסיות מיעוט מוחלטות.

- בנפות המרכזיות, השינויים משמעותיים קטנים יותר (אוכלוסייה יותר גדולה, קשה יותר 'להזיז' את הממוצע, כמצופה). לדוגמא - נפת תל אביב, ירושלים, באר שבע, לעומת צפת, רמת הגולן

e. **תוצרים ומסקנות** (המאגר המחושב מהווה תוצר בפני עצמו, לאחר שיפור ויזואליזציה ע"ב סקאלת צבעים כמפת חום):

i. ישנם הבדלים משמעותיים בין הנפות השונות שלרוב מצביעים על פערים הקשורים בפרפריה (חברתית או גיאוגרפית).

ii. ניתן לראות כי יש מגמה ארצית של שיפור בזכאות לבגרות לפי נפה כאשר באזורים מסויימים חל שיפור ניכר (צפת לדוגמא). לפי נתונים שראינו (ראו פר בית ספר) אכן יש מגמת שיפור, אם זאת ישנם בתי ספר שלא נכנסו לקריטריון זכאות לבגרות ולכן יש להסתכל על הנתונים בעין ביקורתית.

תצלום מסך מתוך `sheet: zacaut, zacaut_per_nafa_mean.xlsx`:

זכאות לבגרות ישראלית לפי נפות ושנים						
	AVERAGE	תשנח	תשנז	תשעו	תשעה	תשעד
חדרה	64.4048	69.8396	68.342	61.8702	60.3447	61.6273
יזרעאל	64.8075	68.9303	67.1938	63.6161	63.6254	60.6717
באר שבע	65.0313	68.5072	66.2173	63.1223	64.5086	62.8011
כנרת	65.8275	66.0632	69.5278	62.2263	67.8438	63.4765
צפת	65.8278	74.75	67.2111	66.3333	62.9556	57.8889
ירושלים	66.5608	71.7829	64.4211	66.2436	65.65	64.7065
השרון	68.1176	72.7488	69.378	67.6775	65.5684	65.2154
עכו	69.2879	69.0035	70.2286	68.8048	69.9675	68.435
אשקלון	69.7399	72.3185	72.04	70.7169	69.2018	64.4224
רמת הגולן	70.9447	78.975	72.7182	73.59	64.1778	65.2625
חיפה	71.7418	73.6833	73.1077	70.6176	70.242	71.0583
פתח תקווה	73.6447	75.0026	74.9644	76.8721	71.1706	70.2136
תל אביב - יפו	73.7006	74.756	74.5574	73.2374	73.3204	72.6317
רמלה	74.2822	79.4667	76.48	74.7525	70.4605	70.2514
רחובות	78.3281	82.2158	80.2482	79.3075	76.3462	73.5229
יהודה ושומרון	78.7966	83.0127	80.815	77.5786	76.5453	76.0314

iii. מבדיקת מגמת השיפור (בין תשע"ד לתשע"ח באחוזים) ניתן לראות מגמות שונות, לחיוב ולשלילה. בהתאם, ניתן להסיק מכך המלצות לקבלת החלטות ומדיניות – תוך העמקה איכותנית ברשויות שהשתפרו משמעותית ובסיבות לכך (דוגמת צפת) וגיבוש מדיניות לסיוע לרשויות בהן אחוז השיפור נמוך (דוגמת עכו).

iv. תצלום מסך מתוך sheet: improvement , zacaut_per_nafa_mean.xlsx :

שיפור בזכאות לבגרות באחוזים מתשע"ד לתשע"ח בחלוקה לנפות		
אחוז זכאות תשנח	שיפור בזכאות לבגרות באחוזים מתשע"ד לתשנח	עכו
69.00348837	0.830698286	עכו
74.7559633	2.924729322	תל אביב - יפו
73.68333333	3.694148	חיפה
66.06315789	4.075033288	כנרת
75.0025974	6.820556927	פתח תקווה
68.50721649	9.085950449	באר שבע
83.01269841	9.182164717	יהודה ושומרון
71.78285714	10.93604665	ירושלים
72.74883721	11.55164941	השרון
82.21578947	11.82335141	רחובות
72.31846154	12.25667789	אשקלון
79.46666667	13.1175099	רמלה
69.83962264	13.325837	חדרה
68.93030303	13.61201499	יזרעאל
78.975	21.01130052	רמת הגולן
74.75	29.12667946	צפת

v. בבחינת הנתונים על ערים (מידע מלא ניתן למצוא בתוצר בקובץ city_mean_zacaut_geo_with_nafa_clean.xlsx = sheet): נראתה מגמה דומה לנפות, של עליה באחוזי הזכאות לבגרות.

vi. בראש טבלת הערים מופיעה במפתיע בית ג'ן – השייכת לנפת עכו (אחת הנפות החלשות במדינה). לצדה ישנן כמובן גם ערים מחתך סוציו-אקונומי גבוה כצפוי (גני תקווה, גבעתיים, השרון, גבעת שמואל, מיתר, שוהם).

- מתוך טבלת זכאות לבגרות לפי ערים:

average	תשנח	תשנו	תשנד	תשנה	תשעד	geo_x	geo_y	Nafa	city	idxCol
31.42	25	31.3	60	15.8	25	35.503111	31.773735	יהודה ושומרון	מגילות ים המלח	מגילות ים המלח
32.13	34.9	34.85	27.15	32.7	31.05	35.35293	32.722572	יזרעאל	עין מאה'ל	עין מאה'ל
38.33	35.775	37.5	29.975	40.53333	47.86667	34.804604	31.246442	באר שבע	אל קסום	אל קסום
38.945	46.65	33.05	39.225	38.5	37.3	34.796603	31.245097	באר שבע	נווה מדבר	נווה מדבר
39.4	58.2	37.9	41.7	30.5	28.7	35.592018	32.96621	צפת	טובא-זנגריה	טובא-זנגריה
41.05	72.7	15.4	56.1	20		35.0418755	31.934401	יהודה ושומרון	מודיעין עילית	מודיעין עילית
41.16	42.53333	43.95	33.26667	45.65	40.4	35.254476	32.921203	עכו	מג'ד אל-כרום	מג'ד אל-כרום
42.36	42.96667	35.33333	39.1	46.13333	48.26667	34.85824133	31.24597	באר שבע	תל שבע	תל שבע
43.22	62.8	36.6	58.1	23.4	35.2	34.912128	32.536655	חדדה	ג'סר אל-זקא	ג'סר אל-זקא
44.08	73.7	28.6	58	30.3	29.8	35.261482	32.717249	יזרעאל	עילוט	עילוט
44.2	58.1	41	50	38.2	33.7	35.1532	32.736287	יזרעאל	בסמת טבעון	בסמת טבעון
45.71619	44.87143	40.67143	51.23333	42.87143	48.93333	35.17177	32.806063	עכו	שפרעם	שפרעם
46.6	48.1	42	39.7	51.7	51.5	35.220571	32.777291	עכו	ביר אל-מכסור	ביר אל-מכסור
46.66	52.6	45.96	44.6	47.14	43	34.6181034	31.311436	באר שבע	אופקים	אופקים
47.57333	50.76667	45.7	57.7	37.65	46.05	34.83931	31.198071	באר שבע	שבג שלום	שבג שלום
48.18	58.05	47.6	57.25	43.7	34.3	35.17799	32.5487265	חדדה	מעלה עירון	מעלה עירון
48.5	59.8	50.9	43.6	40.7	47.5	35.157816	31.804176	ירושלים	אבו גוש	אבו גוש
48.78	51.3	53.6	44.5	44.3	50.2	35.320743	32.682085	יזרעאל	אבסאל	אבסאל
49.34	61.1	56	56.1	41.7	31.8	35.19609	32.690807	יזרעאל	זחור	זחור
49.645	54.825	56.075	51.05	53.075	33.2	35.117063	32.79323	חיפה	זבולון	זבולון
49.68	60.35	55.8	44.3	44.95	43	34.786412	31.758651	אשקלון	יואב	יואב
50.15	59.7	39.2	67.2	44.7	38.3	34.770069	31.917039	רחובות	גן רוה	גן רוה
50.85	49.82	38.7	55.5	56.9	33.9	35.093739	32.970373	צפת	מרום הגליל	מרום הגליל
50.44667	61.25	46.63333	57.65	44.76	41.94	34.7554	31.3928965	באר שבע	הרש	הרש
51.43667	54.96667	54.4	49.86667	46.6	51.35	35.083681	31.249246	באר שבע	כסיפה	כסיפה
51.87	57.85	62.5	65.7	43.3	30	35.0988175	32.7505215	חיפה	רכסים	רכסים
52.09	58.55	59.45	53.55	51.3	37.6	34.761035	31.442291	באר שבע	בני שמעון	בני שמעון
52.19667	74.35	53.63333	60.3	36.7	36	34.981487	32.285791	השרון	קלנסווה	קלנסווה
52.24333	54.86667	45.66667	52.66667	61.15	46.86667	34.950506	32.235466	השרון	טיה	טיה

7	בית אל	בית אל	יהודה ושומרון	31.9416	35.222734	86.05	87.76667	86.86667	86.8	87.76667	87.05
8	אשכול	באר שבע	אשכול	31.302878	34.43135	87.95	80.93333	86.775	87.5	92.9	87.21167
9	אפרת	יהודה ושומרון	יהודה ושומרון	31.65836633	35.15619533	80.93333	86.23333	90.06667	88.36667	93.3	87.78
0	מנשה	מנשה	חדרה	32.467136	35.012976	82.725	87.025	90.3	88.475	94.45	88.595
1	רמת נגב	באר שבע	רמת נגב	31.00441	34.770368	84.9	91.4	79.7	98.9	89.5	88.88
2	בוקעתא	בוקעתא	רמת הגולן	33.201655	35.778757					88.9	88.9
3	רמת השרון	רמת השרון	תל אביב - יפו	32.13915775	34.83804875	84.6	89.725	88.875	88.9	92.525	88.925
4	קצרין	קצרין	רמת הגולן	32.993027	35.689823	89.6	88.3	91.25	91.45	85.75	89.27
5	פקיעין - בוקיעה	פקיעין - בוקיעה	עכו	32.97309	35.326711	92.7	89.9	87.1	87.5	90.4	89.52
6	מזבנת בתיה	מזבנת בתיה	רחובות	31.849076	34.839379	91.4	91.3	91.4	88.6	85.9	89.72
7	שפיר	שפיר	אשקלון	31.697167	34.728724	76.8	91.65	95.93333	90.2	95	89.91667
8	שוהם	שוהם	פתח תקווה	32.003884	34.947397	89.5	87.1	92.1	90.8	91.55	90.21
9	אלקנה	אלקנה	יהודה ושומרון	32.111538	35.0344335	84	85.35	96.6	89.8	97.85	90.72
0	מיתר	מיתר	באר שבע	31.327162	34.941613			91.1		91.9	91.5
1	גזר	גזר	רמלה	31.88859	34.917376	89.76667	90.4	92.63333	94.33333	90.6	91.54667
2	תל מונד	תל מונד	השרון	32.256138	34.921469	92.3	90.1	92.1	96.4	90.8	92.34
3	חורפיש	חורפיש	עכו	33.016688	35.347106	94.8	92	94.9	91	93.4	93.22
4	נס ציונה	נס ציונה	רחובות	31.93143633	34.79887533	91.3	91.9	92.3	95.53333	95.23333	93.25333
5	לב השרון	לב השרון	השרון	32.260652	34.894631	90.9	92	94	94.6	95.5	93.4
6	יבנה	יבנה	רחובות	31.87448625	34.746518	93.05	97.78	89.92	94.6	92.8	93.63
7	דרום השרון	דרום השרון	פתח תקווה	32.133649	34.911086	91.95	92.5	96.35	96.45	91.46667	93.74333
8	כאוכב אבו אל-היג'א' כאוכב אבו אל-היג'א'	כוכב	עכו	32.830676	35.248658		98.3	85.3	96.7	96.3	94.15
9	גדה	גדה	רחובות	31.811044	34.78039175	90.125	92.475	97.05	95.075	96.4	94.225
0	גבעת שמואל	גבעת שמואל	פתח תקווה	32.076281	34.8502555	95.5	92.7	93.5	93.6	97.6	94.58
1	גבעתיים	גבעתיים	תל אביב - יפו	32.0726485	34.80892975	93.8	93.225	96.775	93.275	97.025	94.82
2	גני תקווה	גני תקווה	פתח תקווה	32.059893	34.87532		100	90	100	90	95
3	גוש עציון	גוש עציון	יהודה ושומרון	31.658085	35.118082	92.725	94.7	96.55	94.9	98.375	95.45
4	קריית עקרון	קריית עקרון	רחובות	31.86169	34.822476	94.5	98.2	98.2	93.2	96.1	96.04
5	בית ג'ן	בית ג'ן	עכו	32.964282	35.378468	100	98.5	99	99.5	92.6	97.92

4. **פערים בזכאות לבגרות בתוך ערים:**

a. קובץ קוד: GapPerCityNafa.py

b. **מטרה:** בחינת פערם בתוך עיר בין ביה"ס בעלי זכאות מקסימאלית ומינימאלית, כמדד לאי-שיוויון בתוך רשות מוניציפאלית המחייבת בחינה והעמקה.

0. הסבר התהליך:

i. עבור כל אחת מהשנים המדוברות חישבנו לכל עיר את ההפרש בין בית הספר עם אחוז הזכאות המקסימלי לבין בתי הספר עם אחוז הזכאות המינימלי ושמרנו בקובץ עזר `city_gap_zacaut_with_nafa.xlsx`. לאחר מכן היה צורך בתיקונים מינורים למול בחינה של האקסל, הוספנו עמודת ממוצע ולשם הנוחות הויוזאלית צבענו את הערכים בסקלאת צבעים ומיינו לפי עמודת הממוצע לבחינת דפוסים.

d. תוצר :

i. לאחר מכן איחדנו את הנתונים לקובץ סופי ב city_mean_zacaut_geo_with_nafa_clean.xlsx
sheet =zacaut_gap
שם הוספת עמודת ממוצע וצבענו לפי סקאלת צבעים של מפת חום (מהפער הקטן והטוב לפערים הגדולים). לאחר אפיון התוצאות החלטנו להשמיט ישובים בעלי פער בגודל אפס (לאחר בדיקה מדגמית כי מדובר בישובים בהם יש ביי"ס בודד).

ii. התוצר מונגש גם כשכבה על מפה (שכבה מס' 10).

5. בחינת כמות השקעה בתלמיד בכל בית ספר ובכל עיר :

a. קובץ קוד : Equality.py

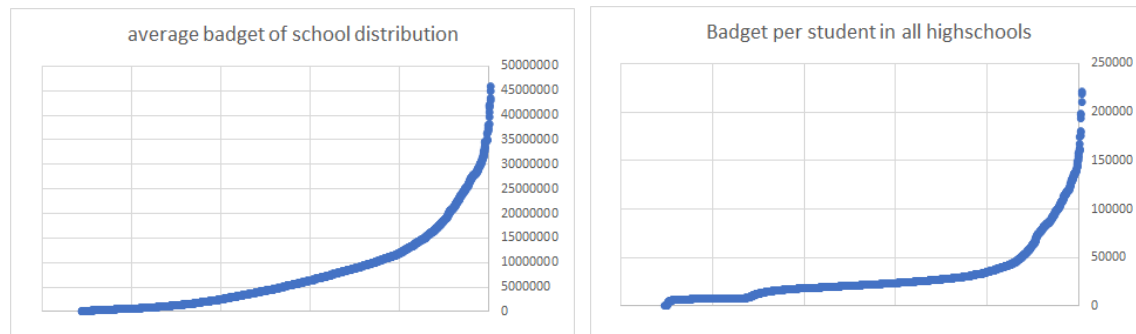
b. מטרה: אפיון הפערים בתקציב לכל תלמיד בבתי הספר תחת ההנחה כי הדבר מהווה פוטנציאל השפעה מרכזי גם על הישגי התלמידים ואיכות הלמידה. ביצענו השוואה ברמת בית ספר ובתוך על עיר.

c. הסבר התהליך :

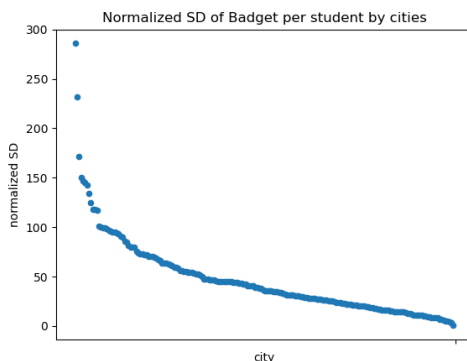
- ביצענו סינון על בתי הספר לאפיון בתי הספר התיכוניים בלבד.
- על כל בית ספר, חישבנו ממוצע תקציב וכן יחס תקציב לכל סטודנט – באמצעות שיקלול נתונים ממספר עמודות.
- לבסוף, יצרנו תצוגה ויזואלית של המידע, וכן יצרנו בסיס ידע שישמש גם ליצרת שכבה למפה.
- ביצענו אווליאציה מדגמית למול קובץ mainBadget_by_shcol.xls בבחינה של מדדים משוקללים, לראות אם סדרי הגודל הגיוניים.

d. תוצר ומסקנות :

- ניתן לראות שביה"ס בהבדלים משמעותיים, שנובעים בחלקם מגודל בתי הספר ומאפייניהם (היקף החינוך המיוחד בביה"ס לדוגמה), שמשפיע למעשה בסופו של דבר על כמות ההשקעה התקציבית עבור כל תלמיד.
- בבחינת בתי הספר, ניתן לראות שההבדלים בין התקציב לתלמיד בכל בית ספר הם במגמה מעריכית. מבדיקה מדגמית של קצוות ההתפלגות, ניתן לראות את ההשפעה של גודל בית הספר על מאפייני התקציב לכל תלמיד ובכך, שחלק מהתקציב של בתי הספר מושקע בעלויות קבועות.
- כמו כן, מבדיקת התפלגות ממוצע התקציב (ללא התחשבות בכמות התלמידים) ניתן לראות מגמה דומה שבה הנתונים מפולגים כמובן באופן שונה.



iv. כמו כן, בבדיקה משוקללת על כל עיר (ע"ב סטיית תקן מנורמלת לסכום התקציב) ניתן לראות כי ישנן ערים עם הבדלים משמעותיים בין בתי הספר שלהן בתקציב לכל תלמיד, דבר שיכול לשמש נתון חשוב עבור הפרט בקבלת החלטות להצטרפות לביי"ס, וכן כסוגייה לבחינה עבור מקבלי ההחלטות.



Normalized SD	City
285.848810	בנימינה גבעת עדה
231.896647	לב השרון
171.361717	כסרא סמיע
150.353198	גבעת זאב
146.979120	הגליל התחתון

5.218368	כפר יונה
4.903493	מזכרת בתיה
4.364335	אשכול
3.116830	ברנר
0.475451	בענה

בתי ספר בעלי תקציבי קיצון (גבוה ביותר ונמוך ביותר לכל תלמיד :

School	School name	City	per student
358127	המגשימים	ירושלים	220848.7
353706	הרים	בנימינה גבעת עדה	218266.5
512491	בנימין רוטמן כדורי	הגליל התחתון	209832.8
620385	גוונים	קריית שמונה	198231.8
731182	פסיפס	מודיעין-מכבים-רעות	197215.4
...			
900126	אורט למינהל טכני	ראשון לציון	149.2963448
378125	אליתים אלערבי	ירושלים	49.51419686
363150	ישיבת פאר יעקב	ירושלים	22.45297428
961011	הילה רטורנו-חסות	מטה יהודה (גבעת שמש)	22.33630769
166538	אורה ושמחה	ירושלים	16.84117647

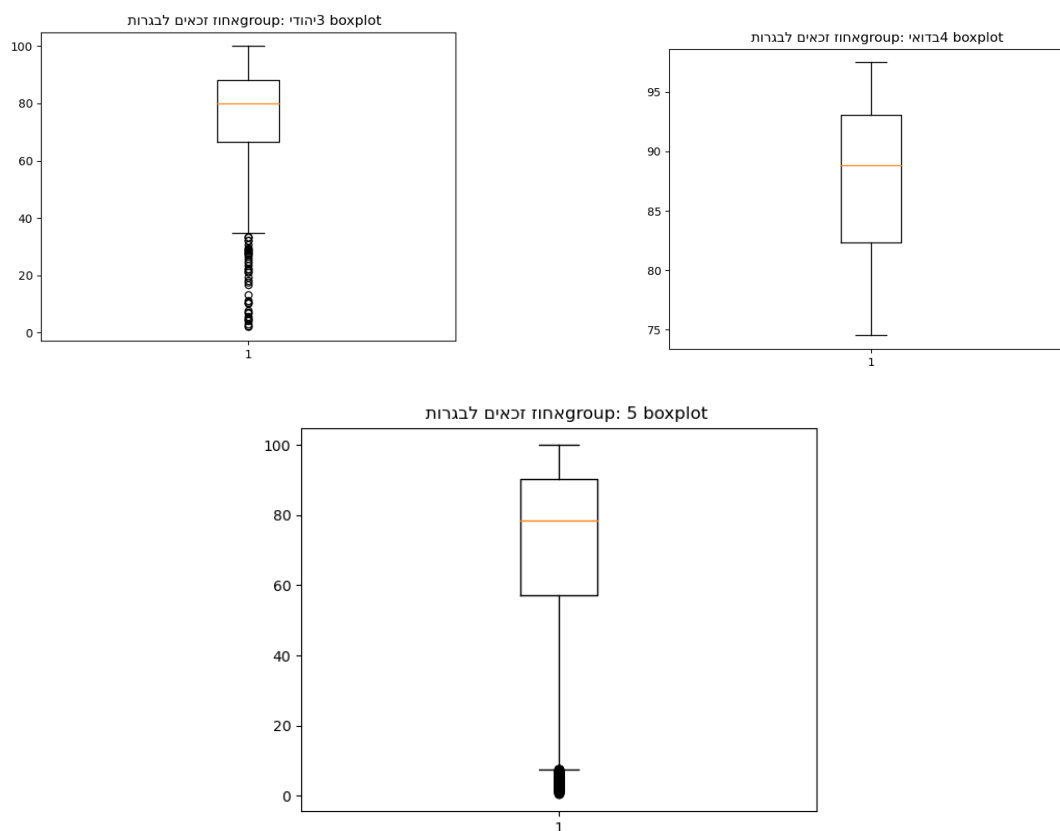
6. בחינת מאפייני חלוקת בתי הספר

- קובץ קוד:** `zacaut.py`
- תצורת עבודה מרכזית:** ניתוח השוואתי ותיאורי של התפלגות הנתונים.
- מטרה:** השוואה בין פילוחים שונים של בתי הספר לתתי קבוצות, למציאת מודלי חלוקה המגדירים בצורה טובה תתי קבוצות, מתוך הבנה כי אופן החלוקה משפיע על תצורת העבודה הרוחבית של מערכת החינוך ודרך כך גם על תקציבים ומשאבים.
- הסבר התהליך:**
 - ע"ב טבלת "school – bagrut.xls" בחנו את ההתפלגות של הזכאויות השונות לבגרות לפי שני מודלים של קבוצת דומים מאותם סדרי גודל (20 במודל שלנו ו27 במודל הקיים), החופפים באופן חלקי:
 - במודל הקיים – כלומר, חילצנו קבוצת דומים מנתוני PROXY שלהם (ממוצעי זכאויות לפי קבוצת דומים).
 - בחלוקה לתתי קבוצות של כל הצטלבות מגזר ואחוזון טיפוח סוציאקונומי.
 - ניתחנו מדדי שונות – כחישוב לפי כל תת קבוצה, ונעזרנו בתוצרים גרפיים של Box Plots, במטרה להציגם ככלי ניהולי לאפיון תתי קבוצות (ממוצעים, ערכי קיצון וכו') ולגיבוש מדיניות תואמת.
 - במקביל, בחנו את אופן התפלגות הנתונים לפי החלוקה שהצענו ביחס לנתונים הקיימים – כהוכחת ערכיות לחלוקה זו עבור חלק מתתי הקבוצות, בניסיון להקטין שונות – כמדד שמעיד על מדיניות אחודה שאינה תואמת את כלל בתי הספר בקבוצה.
- תוצר ומסקנות:**
 - מצאנו כי בבחינת הזכאות ל5 יחידות מתמטיקה, 5 יחידות אנגלית ובגרות בהצטיינות, קיימות תתי קבוצות בעלות שונות נמוכה יותר (לדוגמה בקבוצת אחוזון 1 במגזר הערבי החופפת ברובה לתת קבוצה קיימת):



חלוקה	zacaut kind	mean	Var
קיימת	אחוז זכאות 5 יחידות אנגלית	45.94624	410.9903
1 ערבי	אחוז זכאות 5 יחידות אנגלית	45.775	347.6292

ii. בקבוצה קיימת המכילה ערבוב בין אחוזון 3 היהודי לבין אחוזון 4 הבדואי, יש פער משמעותי בין הישגי בתי הספר היהודים לאלו הבדואיים – כלומר ניתן לראות את "המרוויח והמפסיד" מאותה תת חלוקה, ככלי לקבלת החלטות ניהוליות. (להלן ה-BoxPlots של תתי הקבוצות שלנו, השייכות לאותה "קבוצת דומים" קיימת (מס' 5) המופיעה בגרף מתחת.



iii. מבדיקת מקרה בוחן, להשוואה נקודתית בין קבוצות דומים קיימות לבין הנתונים המקבילים להן בקבוצות החלוקה שלנו, כך לדוגמה: עבור קבוצת הדומים בה ממוצע הזכאות לבגרות הינו 56.9 עולה כי היא מאופיינת כולה באחוזון [5'] ובמגזרים ['יהודי', 'בדואי', 'ערבי', 'דרוזי'], וככזו-השווה בה הינה: 356.192330881591. בהשוואה לחלוקה שלנו, ניתן לראות בפירוש שתת קבוצה זו מתפלגת בצורה לא אחידה הן בשונות תתי הקבוצות והן בממוצעים, כך שהיא עשויה לגרום לפגיעה בבתי ספר בשולי החלוקה.

quintile	sector	zacaut kind	Mean	Var
5	יהודי	אחוז לבגרות זכאים	60.00433	569.4053
5	בדואי	אחוז לבגרות זכאים	50.46413	220.1019
5	ערבי	אחוז לבגרות זכאים	60.66691	420.9802
5	דרוזי	אחוז לבגרות זכאים	77.58864	108.9801

תוצר נתוני הזכאות לבגרות בהשוואה בין החלוקה לפי תתי קבוצות מגזר ואחוזון טיפוח לעומת החלוקה הקיימת :

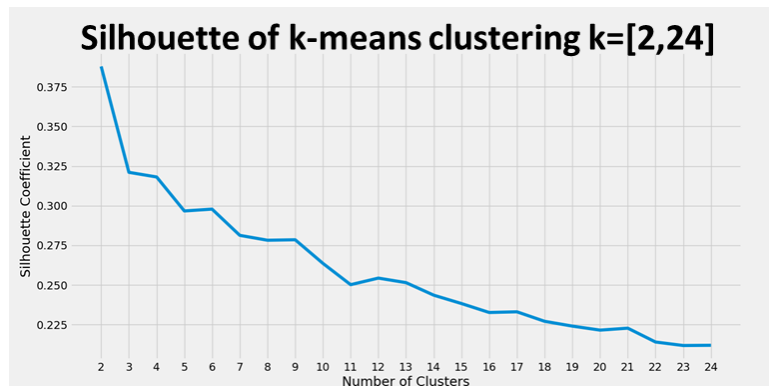
אחוזון טיפוח ומגזר				חלוקה קיימת ל"קבוצות דומים"		
quintile	sector	Mean	Var	#	mean	var
5	יהודי	60.00433	569.4053	1	22.40633	378.0855
5	בדואי	50.46413	220.1019	2	25.38506	513.5985
5	ערבי	60.66691	420.9802	3	26.744	470.2158
5	דרוזי	77.58864	108.9801	4	31.92857	526.9009
נוער בסיכון	יהודי	30.6102	524.8423	5	32.42593	482.3055
נוער בסיכון	בדואי	23.6	246.42	6	56.02778	447.7768
נוער בסיכון	ערבי	17.34405	184.3697	7	56.355	486.1028
נוער בסיכון	דרוזי	4.2	-	8	58.25743	405.4378
4	יהודי	68.58438	454.6857	9	60.65068	401.991
4	בדואי	87.4	73.376	10	62.9973	428.215
4	ערבי	68.87017	337.9727	11	64.37871	446.2432
4	דרוזי	79.33243	196.1278	12	67.49049	459.341
3	יהודי	74.51467	423.9713	13	71.40598	411.1913
3	ערבי	80.07222	176.2229	14	71.67958	464.4517
3	דרוזי	69.2	-	15	74.42667	293.772
2	יהודי	80.01541	343.9218	16	74.61277	304.5173
2	ערבי	91.43714	78.09512	17	74.85419	371.2916
2	דרוזי	99.1	1.73	18	76.56447	466.6575
1	יהודי	86.73019	239.8388	19	77.37645	382.588
1	ערבי	95.375	52.61583	20	80.44337	381.1697
				21	82.75329	279.5271
				22	84.11098	230.0887
				23	85.61726	253.4089
				24	86.12049	237.6285
				25	86.24455	284.6688
				26	86.79561	236.8335
				27	88.8963	184.3151

7. קלסטרינג לפי אחוזי בגרויות

- a. קובץ קוד: cluster for schools.py
 b. תצורת עבודה מרכזית: שימוש בקלאסטרים במודל K-Means.
 c. מטרה: השוואה בראי התוצאה – כלומר, השוואה בין בתי ספר מרקעים שונים החולקים את אותו מרחב הישגים, וכן לבחון למול חלוקות למגזרים ואחוזוני טיפוח.
 d. הסבר התהליך:

i. התבססנו על k-means מתוך ניסיון לבחון את פילוח הקבוצות בסדרי גודל ידועים, מתוך הנחה שיישום של השיטה בפועל ידרוש תפעולית עבודה במספר קבוצות בסדר גודל מסוים. את הקלאסטרים יצרנו על בסיס כל נתוני הזכאות לבגרות השונים (כללי, יח"ל במתמטיקה ואנגלית, הצטיינות וכו') ובכך ייצרנו תתי-קבוצות בעלי מאפיינים יחודיים שלא ניתן לחלק באמצעות פילוח קטגוריאל פשוט.

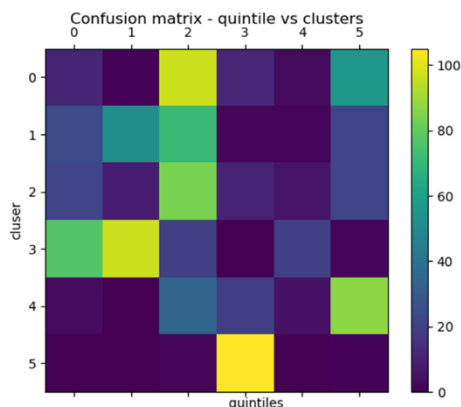
ii. בהתאם, לאחר בדיקה גדלי K אפקטיביים ע"ב SSE (שלא הניב תוצאה מובהקת) ועל Silhouette, שממנו הפקנו יותר תועלת, ובהתאם התמקדנו ב-3 חלוקות: {6,17,21} ובחינת התפלגותם למאפייני בתי הספר (מגזר, אחוזוני טיפוח, סוג פיקוח וכו') ע"ב טבלת "school – bagrut.xls".

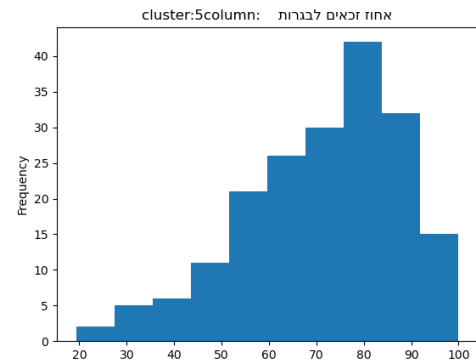
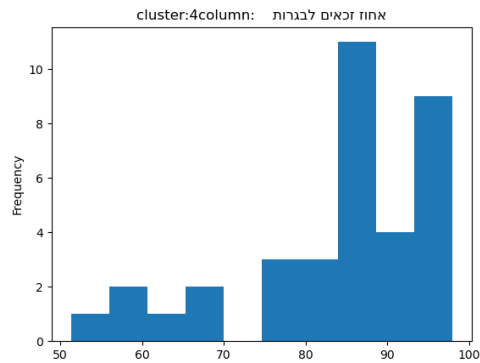
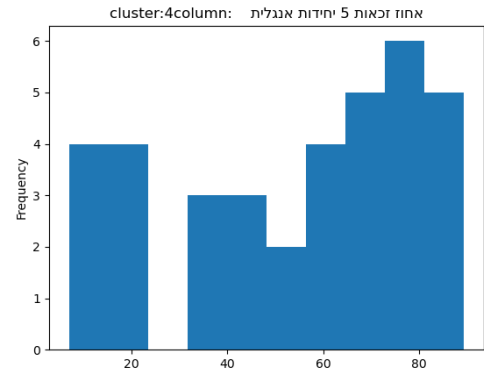
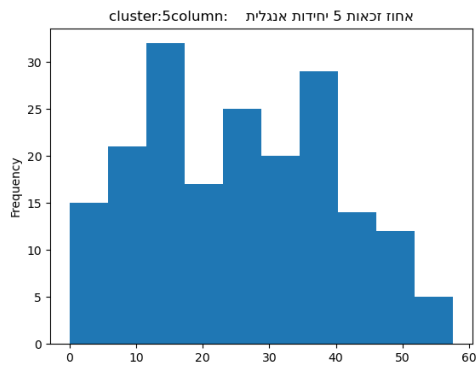


iii. נעזרנו בכלי Confusion Matrix - להשוואה בין הקלאסטרים לבין חלוקות מאותם סדרי גודל, לצד פלט סטטיסטיקה תיאורית על כל קלאסטר וכל תצוגה בהיסטוגרמות של מאפייני בתי הספר בכל קלאסטר.

e. תוצר ומסקנות:

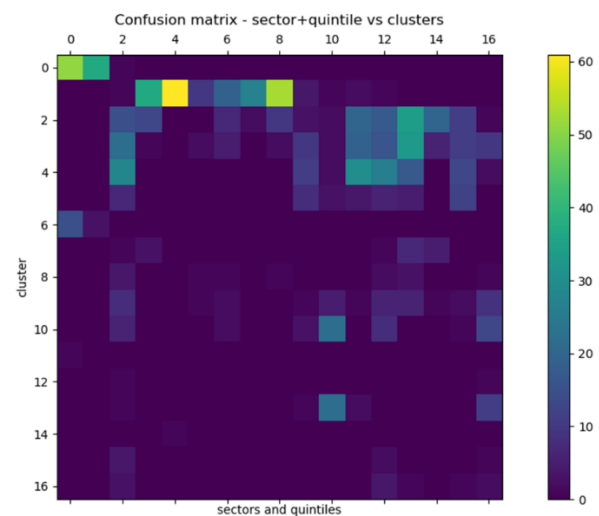
i. כך לדוגמה, בבחינת החלוקה ל-6 קלאסטרים, השוואנו את הקלאסטרים לאחוזוני הטיפוח (הממוספרים 1-5, וכן אחוזון נוער בסיכון סומך כ-0), וניתן לראות כי אחוזון טיפוח 3 מחולק בין רוב המצוי בקלאסטר 5 לבין מיעוט הנמצא בקלאסטר 4 (ומיעוט בנוספים). כך, בהסתכלות על הישגיהם השונים – ניתן להבנות תוכניות יחודיות לבתי ספר מאותו אחוזון טיפוח בהתאם לקבוצת השווים להם בהישגים – כך לדוגמה ניתן לראות את ההתפלגות השונה בזכאות ל-5 יחידות אנגלית, וכן בהתפלגות הזכאות לבגרות באופן כללי.





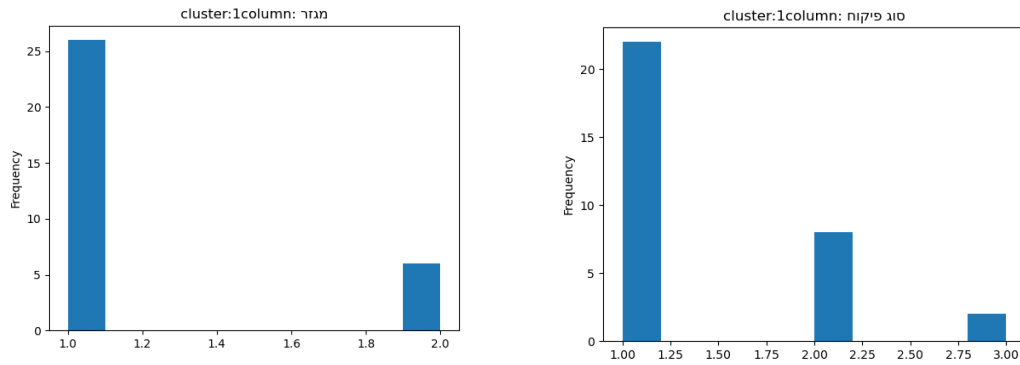
ii. דוגמה נוספת לשימוש זה ניתן לראות גם בהשוואה בין החלוקה שהגדרנו מעלה המשלבת מגזר ואחוזון טיפוח (לדוגמה: "אחוזון טיפוח 3 מגזר יהודי"), לקלסטרים של הזכאות לבגרות (מאותו סדר גודל). כך לדוגמה, בבחינת החלוקה ל-6 קלאסטרים, השוואנו את הקלאסטרים לאחוזוני הטיפוח (הממוספרים 1-5, וכן אחוזון נוער בסיכון סומן כ-0), וניתן לראות כי אחוזון טיפוח 3 מחולק בין רוב המצוי בקלאסטר 5 לבין מיעוט הנמצא בקלאסטר 4 (ומיעוט בנוספים). כך, בהסתכלות על הישגיהם השונים – ניתן להבנות תוכניות יחודיות לבתי ספר מאותו אחוזון טיפוח בהתאם לקבוצת השווים להם בהישגים – כך לדוגמה ניתן לראות את ההתפלגות השונה בזכאות 5 יחידות אנגלית, וכן בהתפלגות הזכאות לבגרות באופן כללי.

מקרא	
סימון	קבוצה
0	יהודי נוער בסיכון
1	יהודי אחוזון 1
2	יהודי אחוזון 2
3	יהודי אחוזון 3
4	יהודי אחוזון 4
5	יהודי אחוזון 5
6	ערבי נוער בסיכון
7	ערבי אחוזון 2
8	ערבי אחוזון 3
9	ערבי אחוזון 4
10	ערבי אחוזון 5
11	בדואי נוער בסיכון
12	בדואי אחוזון 4
13	בדואי אחוזון 5
14	דרוזי אחוזון 2
15	דרוזי אחוזון 4
16	דרוזי אחוזון 5



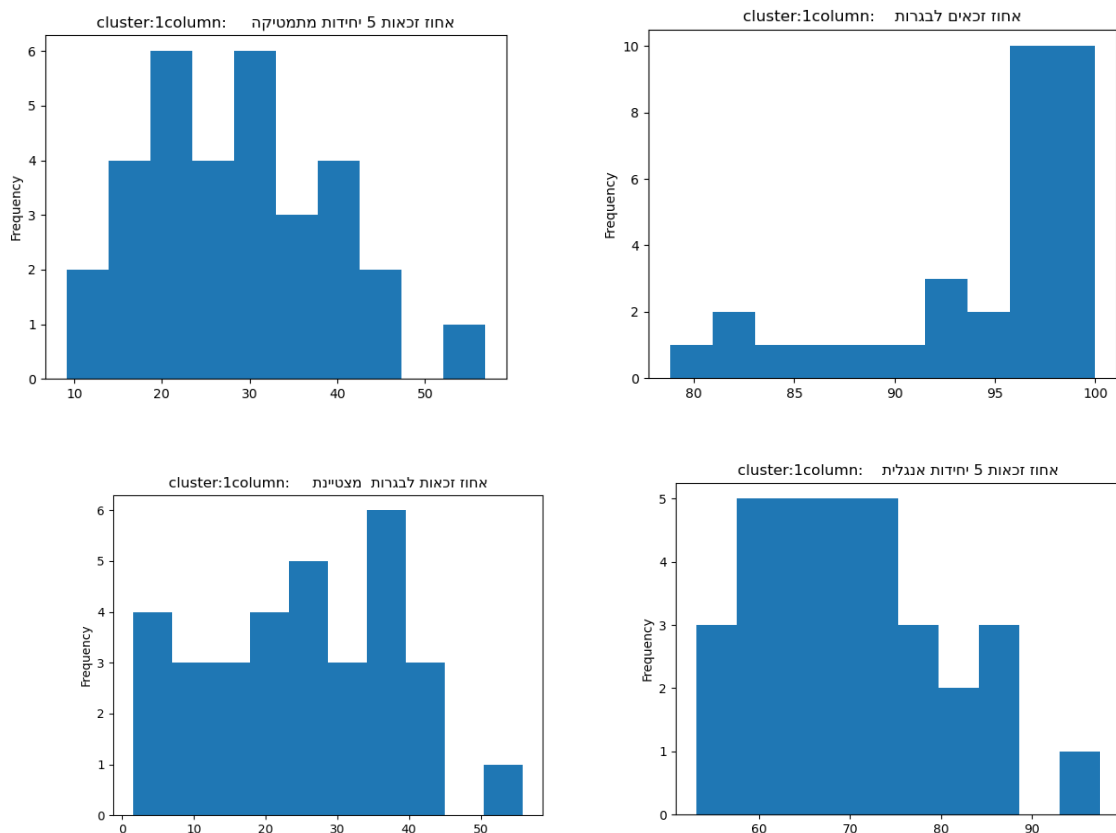
.iii

ניתן לאפיין זיקה ברורה בחלק מהקבוצות למאפייני הישגים מסוימים, לדוגמה ניתן לראות כי בתי ספר יהודים מאחוזונים 3 ומטה חולקים את אותו קלאסטר הישגים עם בתי ספר ערבים באחוזונים 2-3, דבר שיאפשר שיתופי פעולה או פעילות של מערכת החינוך שלא היתה מתגבשת. מהעמקה במקרה מבחן זה ניתן לראות כי מרבית בתי הספר המשוויכים לקלאסטר זה הם מהמגזר היהודי (כ-83%) וברובם הם בפיקוח ממלכתי דתי (סימול 1) המהווים כ-80% מהקלאסטר.



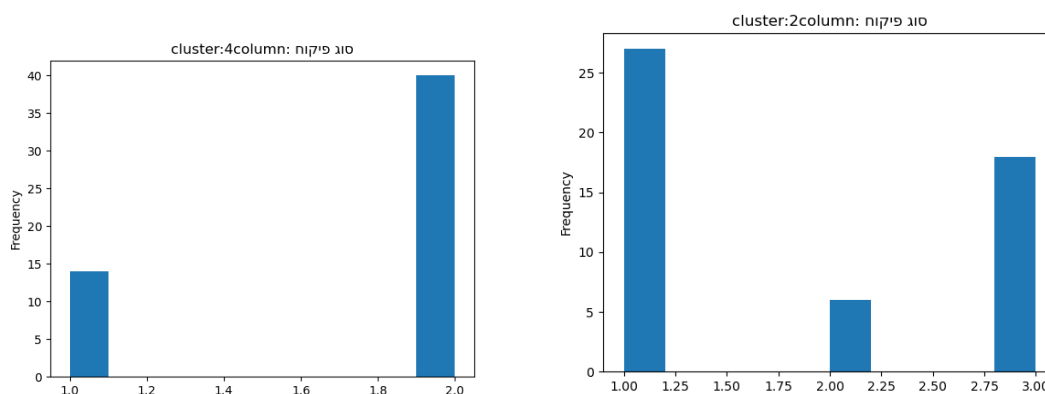
.iv

בבחינת ההישגים המאפיינים את הקלאסטר, עולים אחוזי זכאות גבוהים לבגרות בכלל, ואחוזים גבוהים מהממוצע בפרט בבגרות בהצטיינות וב-5 יח"ל מתמטיקה ואנגלית. כלומר, מדובר בקבוצת בתי ספר חוצת מגזרים וסוג פיקוח (ערבי-ממלכתי לעומת ממלכתי-דתי יהודי).

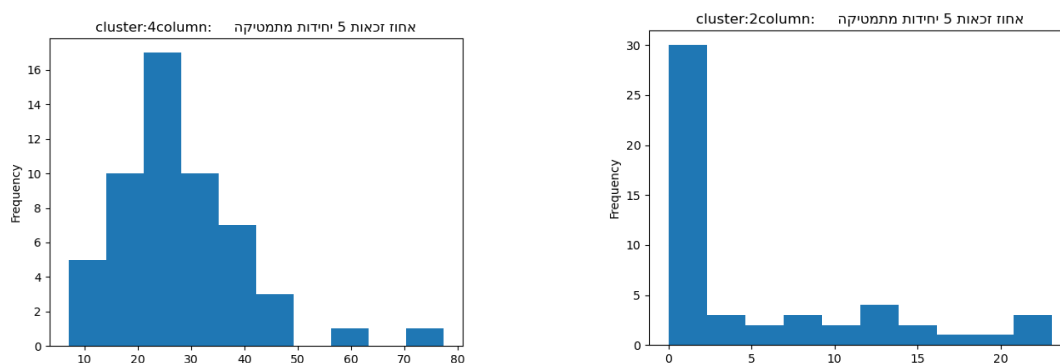


v. כמו כן, ניתן לראות כי מרבית בתי הספר הבדואים (קבוצות 11-13) מאופיינים באותם קלאסטרי הישגים בסיסיים עם בתי ספר יהודים מאחוזון טיפוח 2 אשר מבחינת המקרה עולה כי מרביתם ממלכתי דתי וחרדי. זאת כל הנראה בין היתר לאור פערים דומים בזכאות ל-5 יחידות במתמטיקה. כלומר ניתן לגבש מדיניות לקבוצה משותפת שלא חולקת כלל נתונים סוציאקונומיים ומגוריים.

מאפייני בתי הספר בקלאסטרים 2 ו-4 :



זכאות לבגרות ב-5 יח"ל במתמטיקה :



תוצרים אינטראקטיביים :

כפי שציינו מעלה, המרכיב השני בפרויקט הוא יצירת ממשקים אינטראקטיביים לניתוח והעמקה בתוצרים ובנתונים המעובדים.

1. מפות - גוגל מפת

a. תצורת עבודה מרכזית: יצירת שכבות מידע ע"ב מפה.

b. קישור לתוצר:

i. שכבות המידע הגיאוגרפיות בממשק Google maps
https://www.google.com/maps/d/edit?mid=1OWSj1btTp1KNaz1a5FFvL_bEGC_yyHv0r&usp=sharing

ii. סרטון הדגמה - <https://www.youtube.com/watch?v=Ro5ZRn4kIyk>

c. מטרה: שימוש בניתוחי הנתונים שביצענו ובמידע מנוקח, והנגשתם כשכבות לניתוח אקספלורטיבי ע"ב מפה.

d. הסבר על התהליך והתוצר:

i. יצרנו שני סוגים של שכבות מידע: שכבות הנגשת מידע רקע כללי – לצרכי סקירה והשוואה שכבות המידע המסוכסם (שכבות 1-7); שכבות מבוססות ניתוח ונתונים אגרגטיביים בבחינת בתי ספר וערים (שכבות 8-10). בכך:

א. שכבה 1: תיכונים לפי נפה – ניתוח שייכות ביי"ס לנפה אזורית והצגה בחלוקה לצבעים

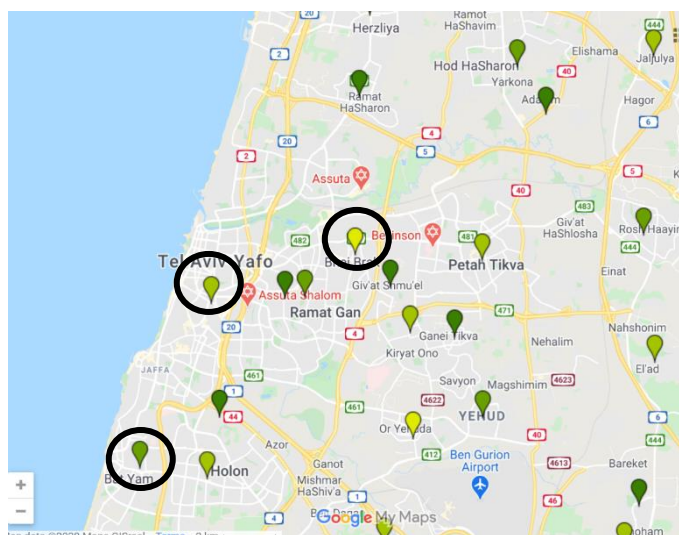
(ע"ב קובץ מעובד - schools_data_only_hs.xlsx).

- ב. שכבות 2-3 : תיכונים לפי מגזר (יהודי, ערבי, דרוזי ובדואי) וסוג פיקוח (ממלכתי, ממל"ד, חרדי), בהתאמה. (ע"ב קובץ מעובד - school_zacaut_nafa_jewish.xlsx).
- ג. שכבות 4-5 : זכאות ל-5 יח"ל במתמטיקה ובאנגלית, בהתאמה, והצגת בתי הספר לפי סקאלת חוס (גיוני כחול מבהיר לכהה) (ע"ב קובץ - zacaut_tsah_full.xlsx).
- ד. שכבות 6-7 : בתי ספר לפי היקף תקציב לכל תלמיד בתשע"ח, והיקף שעות לימוד פרטניות בתשע"ח, בהתאמה. בתי הספר צבועים לפי מפת חוס (מבהיר לכהה). (ע"ב ניתוח מעובד של נתוני התקציב המובע בקובץ - per_stud_data_for_map.xls).
- ה. שכבות 8-9 : ממוצעי זכאות כללית בכל עיר לשנת תשע"ח, וממוצע זכאות בערים בין השנים תשע"ד לתשע"ח, בהתאמה. צבועים כמפת חוס (ירוק-אדום). (ע"ב קובץ מעובד - city_mean_zacaut_geo_with_nafa_clean.xlsx).
- ו. שכבה 10 : מיפוי פערי זכאות לבגרות בין בתי ספר קיצוניים (מיני מול מקסי) באותה העיר בשנת תשע"ח. הערים צבועות ע"ב סקאלת מפת חוס הפוכה (אדום – פער גדול, עד ירוק – פער נמוך). (ע"ב הקובץ בסעיף ה').

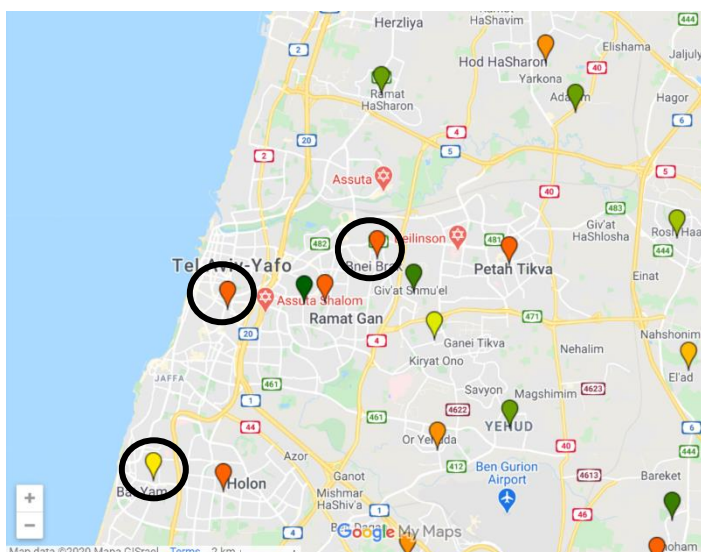
e. מסקנות מהמפות:

- i. הבדלים בין ערים סמוכות – בבחינת ערים סמוכות בעלות פערי זכאות, ניתן לראות הבדלים גם בהיקף שעות הלימוד הפרטניות. כך לדוגמה: בהסתכלות על בת ים, תל אביב ובני ברק, ניתן לראות כי הממוצע הגבוה ביותר הינו דווקא בבתי ים, לאחריו תל אביב ולבסוף בני ברק, והפערים בין בתי הספר בעיר גדלים באותו סדר בהתאמה והיקף שעות הלימוד הפרטניות גדל בהתאמה.

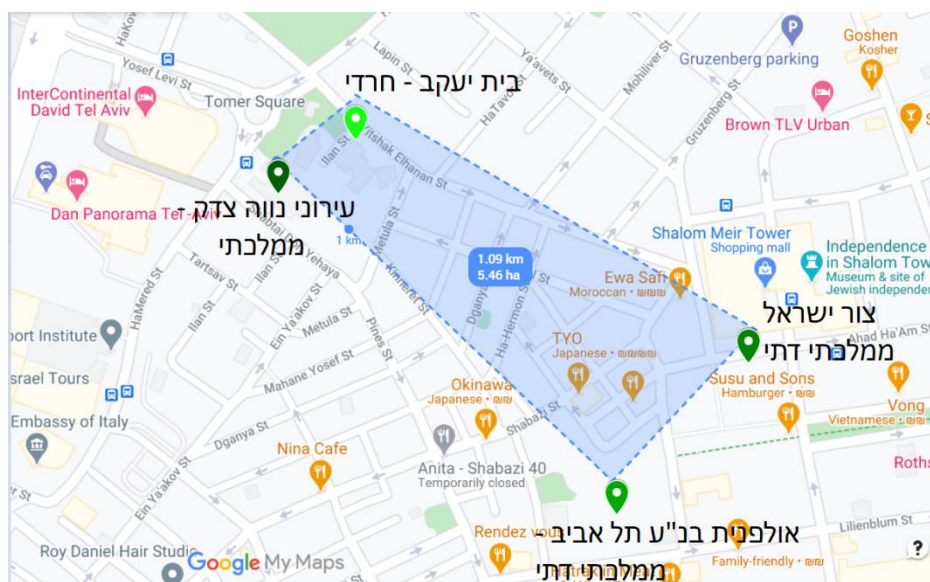
מפת ממוצעי הבגרות:



מפת הפערים בזכאות בין בתי ספר בעיר:



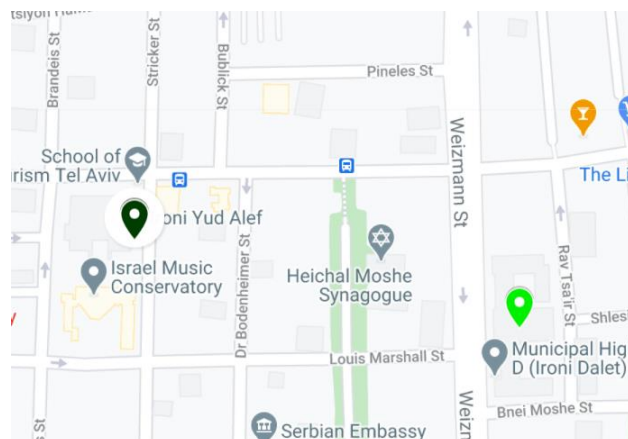
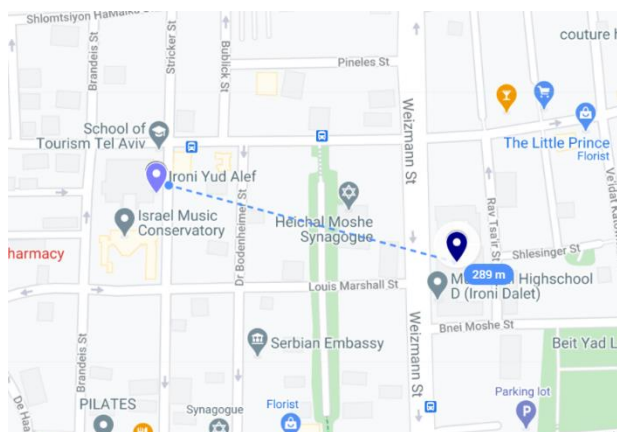
ii. הבדלים בתקצוב באותה עיר בין בתי ספר סמוכים – ע"ב המפה, בולטים הבדלים משמעותיים בתקציב לכל תלמיד בין בתי ספר סמוכים. כך לדוגמה, בתל אביב – מבדיקה על השכבות השונות, בולט כי בתי הספר ממערכת פיקוח שונה, ההבדלים בזכאות לבגרות מקבילים להבדלים בתקציב וכן להבדלים בהיקף שעות לימוד פרטניות (המושפע גם מהיות אחד בתי הספר מותאם לחינוך מיוחד). מקרה מנוגד, בולט בין תיכון י"א ועירוני ד' הסמוכים, בהם בולטים פערים משמעותיים בתקציב ובשעות הפרטניות במגמות הפוכות (מהעמקה- תיכון י"א מיועד לנוער מרקע קשה).



השוואה בין תיכון י"א לתיכון עירוני ד' (מרחק של 289 מ'):

מפת שעות פרטניות

מפת תקצוב

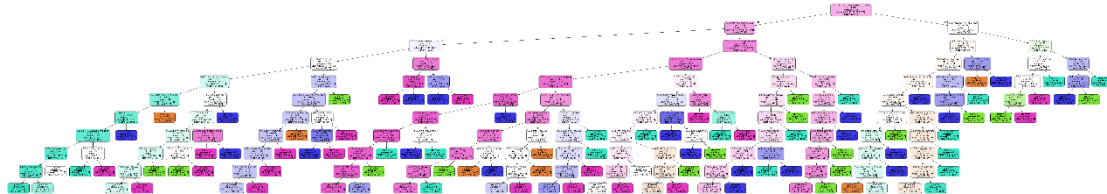


2. מודל לומד אקספלינאבילי כעזר למדיניות

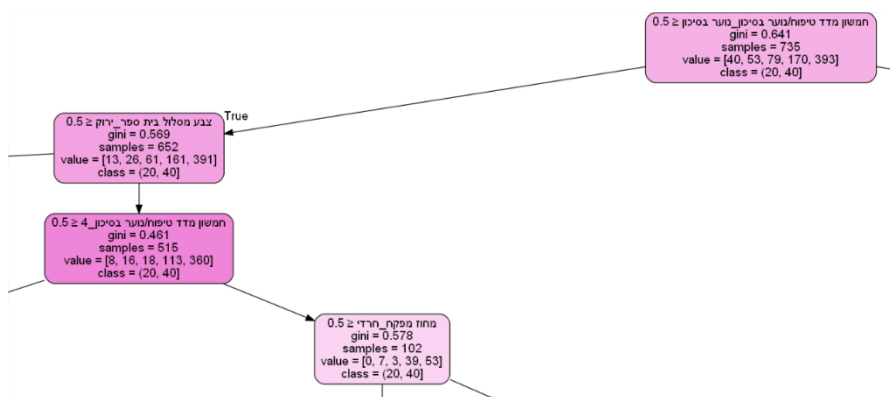
- יצירת עבודה מרכזית: יצירת ממשק להפקת עץ החלטות.
- קישור לתוצר:
 - קובץ פיית'ון: DesTree (ע"ב קובץ גולמי school – bagrut.xlsx).
 - סרטון הדגמה - <https://youtu.be/dghC7nIH-ts>
- מטרה: ניתוח מאפייני התפלגות הזכאות לבגרות של בתי הספר, ממוקד ביעד זכאות אחד, ונועד לייצר "מפת דרכים" להבנת הגורמים הקורלטיביים להישגים שונים.
- הסבר על התהליך והתוצר:

- a. בחרנו להשתמש בעץ החלטות, כמודל הניתן להבנה (אקספלינאבילי) הלומד על המידע בשנה בודדת ומפלג את בתי הספר לפי נתוני השייכות שלהם לרשויות, מגזרים וצורות פיקוח שונות.
- b. המודל מונגש באמצעות ממשק משתמש המקבל את הנתונים הרצויים – סוג זכאות רלוונטי, שנה לבחינת הנתונים, ומאפייני המדידה- תוך מתן אפשרות לשאול שאלה על ציון סף (לדוגמה – 60% ומעלה) או לפי חלוקה לתתי-טווחים קטגוריאליים (לדוגמה – 0-20, 21-40 וכו'). המודל לומד על סט לאימון ומנגיש את רמת האמינות שלו (בבדיקה על סט נתונים נפרד לבחינה) ומציג לבסוף תוצר ויזואלי של עץ החלטות עם פרטי התנאים המבדלים. כך, מקבל החלטות יכול להשתמש בפרמטרים האלו כדי למצוא מוקדים מרכזיים של פערים (פיצורים המבדלים באופן מרכזי בין הליבלים השונים של העץ) ולדייק את מדיניותו למול פערים אלו.
- c. בחרנו לחלק את הדאטה בחלוקה של 75% לאימון, מתוך הנחה כי המודל משמש בעיקר לניתוח המציאות ולא לניבוי בפועל, והדאטה לבחינת התוצאות נועד לשמש כבדיקת מציאות של המודל שהשתמש יוצר ולא מעבר לכך.
- e. **מסקנה לדוגמה:**

- a. לאחר הפקת עץ זכאות כללית לשנת תשע"ח, בחלוקה לקבוצות טווח בקבוצות של 20%, התעמקנו בראש העץ בענפו השמאלי.
- b. בולט כי שייכות לקבוצת נוער בסיכון היא הממד המשמעותי ביותר לפילוח, ומבין בתי הספר העונים לקריטריון זה, שייכות בייס "למסלול ירוק" (מדד של מערכת החינוך) היא המבחן המשמעותי אחריו.
- c. כך יוכל האחראי במערכת החינוך להתמקד בבעיות המרכזיות בתחומו בהתאם למאפיינים השונים של בתי ספר – כלומר להתמקד בתת חלוקה המגדירה באופן משמעותי ביותר את האימפקט על שייכות לטווח מסוים של זכאות לבגרות אותו רוצים לתקן או שאליו רוצים לשאוף.



העמקה בראש השמאלי של העץ:



מכשולים -

במהלך העבודה נתקלנו בכמה מכשולים מרכזיים.

1. **ריבוי ופיזור המידע** - המידע שעמד לרשותנו היה נתון בקבצים רבים, כל קובץ עם מדדים שונים, חלקים ברמה הבית ספרית וחלקים ברמה העירונית. בשלב הראשון, לפני שהתחלנו לנתח את המידע היה עלינו לסדר את אותו. יצרנו כמה קבצי בסיס המכילים את רוב הנתונים הרלוונטים (כל קובץ מורכב ממספר קבצים).
2. **התמודדות עם פערים במידע** - למרות שכל המקורות שהיו בידינו נלקחו ממקורות רשמיים, להפתעתנו הייתה חפיפה חלקית בלבד בין הקבצים השונים. התחלנו את הפרויקט כשאנו בוחנים קרוב לעשרת אלפים שורות (כל שורה מייצגת בית ספר בשנה מסוימת). התצמצמו לכמעט אלף בתי ספר רלוונטים (תיכונים) שעליהם היה לנו מידע בקבצים השונים. היה צורך בתיקונים ובדיקה ידנית, שכן הבדלים ברישום (שגיאות כתיב, שמות מתורגמים מערבית, כתיב מלא וחסר, התמודדות עם רווחים) גרמו לפערים בטבלאות המאוחדות.
3. **מציאת הכלים המתאימים לניתוח המידע והצגתו** - אחד הקשיים המרכזיים שלנו היה להבין מה הכלים המתאימים לניתוח המידע. היה עלינו ללמוד אלגוריתמים לומדים, אלגוריתמים של קלסטור, שימוש ב API של גוגל מפות וספריות שונות לשליפת מיקומים (להשלמת מידע חסר).
4. **התאמה של מידע קטגורי לניתוחים סטטיסטיים כמותיים/ למודלים של MACHINE LEARNING** - בקלאסטרניג היה צריך לסדר את המידע על מנת שהאלגוריתם יוכל לקרוא את הנתונים ולהשוות בין פילוח קטגורי (לדוגמה לפי מגזרים) לעומת פילוח לפי נתונים רציפים (לדוגמה אחוז זכאות לבגרות).

- בעץ החלטות - הפיכת נתון קטגורי לבינארי על ידי ייצור מאפייני Dummie והפיכת נתונים קטגוריים לרציפים בדומה לקלסטרינג (לדוגמא עץ ההחלטות שלנו מתבסס אומנם על כ 8 עמודות אך על יותר מ200 פיצ'רים, היות וכמעט על לכל עמודה ייצרנו מאפיין Dummie).

עבודה אפשרית עתידית -

הפרויקט מהווה הוכחת היתכנות לשימושים קבועים במידע הזה בצורה נוחה וויזואלית, בהתאם - עבודה עתידית מרכזית יכולה להיות על נתונים עדכניים מהשנים האחרונות (2018-2020) ובחינת המודלים גם עליהם. כמו כן, ניתן לבצע העמקות נוספות הממוקדות במאפייני תקציב כרכיב הנמדד - לבחון קלאסטרים ועץ החלטות על בסיסים (כלומר להתמקד בנתוני המוצא של בתי הספר ולא בהישגים כרכיב הנמדד). תכליות אפשריות של הפרויקט הינם: כבסיס לגורמים במערכת החינוך לצורך קבלת החלטות בתקצוב והעברת משאבים, בהסתמך על הנתונים והתוצאות מהשנים האחרונות, מבט חיצוני על המערכת ותפקודה. כמו כן, כמנגנון הערכה של בתי ספר, ערים ונפות לאיתור הצטיינות, מגמות ופערים הדורשים מענה. בנוסף, ברמת בית הספר הדבר מייצר בסיס לשת"פ ולמידה הדדית החוצה מגזרים, עם בתי ספר מקבילים ברמת הקשיים או ההישגים. ברקע, עבור הורים המתלבטים היכן להשתקע או תלמידים שמתלבטים לאיזה תיכון ללמוד, יכולים לאתר את העיר והתיכון המתאים להם בהתאם למאפיינים ספציפיים.

סיכום כללי -

בפרויקט לקחנו מאגרי מידע שונים והתנסנו במודלים וכלים שונים לניתוח נתונים, תוך התמודדות עם פערי מידע ונתונים חסרים, פערים בין מקורות שונים. הפרויקט היה מלמד מאוד על הכלים והמימושים שלהם ותוצאותיו משקפות את החשיבות בניתוח והעמקה בנתונים כשלב בסיס ובהצגה אפקטיבית ונוחה של המסקנות העולות מהן. קונקרטית, השימוש במודלים שונים לאפיון וחלוקה של בתי הספר ואיתור קורלציות ומאפיינים ברורים בסאב-סטים האלו, מוכיח את השימושיות של ניהול ומדיניות מוכוונת נתונים ככלים שיאפשרו הבנה של מציאות מורכבת ודיוק של תהליכי עבודה.